Lecture 06

Jose Capco (jcapco@risc.jku.at)

Method 1 (Fermat's Factorization): An odd number n can be factorized iff n is the difference of squares of two non-consecutive integers. So if n is an odd composite number then there is $x, a \in \mathbb{N}$ such that $n = (x-y)(x+y)$. If we suppose $n = pq$ for non-trivial $p, q \in 2\mathbb{N}+1$ then we can choose $x = \frac{p+q}{2}$ and $y = \frac{p-q}{2}$. Fermat's factorization algorithm starts from $x_0 = \lceil \sqrt{n} \rceil$ and iterates as follows

Algorithm 1: Fermat's factorization algorithm

Input: A composite (odd) number $n \in \mathbb{N}$
Output: A proper integer factors of n

Set $x = \lceil \sqrt{n} \rceil$
while $x^2 - n$ *not a square* **do**
  Set $x = x + 1$
end
return $x \pm \sqrt{x^2 - n}$

For a composite odd n, there is a k such that $(x_0 + k)^2 - n$ is a positive integer and so $(x_0 + k) \pm \sqrt{(x_0 + k)^2 - n}$ will be a non-trivial factor of n. This method works very well if n has two factors that are not too far away from each other. In fact, a rule of thumb is that one uses Fermat's method if $|p - q| \leq n^{1/4}$.

Exercise 1. Suppose $n = pq$ is an odd composite number and that $|p - q| < n^{1/4}$ as in the above procedure. Find the maximum number of steps in Fermat factorization method that is needed to obtain p and q.

A serious improvement to the above procedure is the following ...

Method 2 (Quadratic Sieve): This procedure was invented by Pomerance. Here is the procedure, given an odd composite $n \in 2\mathbb{N} + 1$ define $x_0 = \lceil \sqrt{n} \rceil$ and ...

1.) Choose a factor base $p_1, \ldots, p_k \in \mathbb{P} \cup \{-1\}$, say prime numbers less than some positive $B \in \mathbb{R}$. -1 should be included in the factor base.

2.) Find sequences $v_1, \ldots, v_l \in \mathbb{Z}$ that are factored by these primes, e.g. B-smooth numbers, such that

   i. The v_i's represent numbers of the form $(x_0 + r_i)^2 \bmod n$, where $r_i \in \mathbb{Z}$ is in some small interval, say $[-n^{1/4}, n^{1/4}]$

   ii. The vectors $\vec{e}_1, \ldots, \vec{e}_l \in \mathbb{F}_2^k$ defined by

$$\vec{e}_i := (v_{p_j}(v_i) : j = 1, \ldots, k)$$

   are linearly independent in $\mathbb{F}_2^k$. It could be that one of the $\vec{e}_i$ is $\vec{0}$, in this case we are actually doing Fermat's factorization

3.) We can find a kernel of $(\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_l) \in (\mathbb{F}_2)^{k \times l}$ and thus find product of some of the v_i's that is a square number. Without loss of generality let us assume they are the product of all of the $v_1, \ldots, v_l$ (see matrix below).

4.) So there is an $x, y \in \mathbb{N}$ such that $x^2 = \prod_{i=1}^{l} (x_0 + r_i)^2$ and $y^2 = \prod_{i=1}^{l} v_i$ and the hope is that $\gcd(x \pm y, n)$ is a proper non-trivial divisor of n.

Example 1. Quadratic sieve is best learned by looking at an example consider $n = 87463$ and consider the v_i chosen such that they are 29-smooth. These are shown in the table blow

Then the quadratic sieve matrix is as shown

$$\begin{pmatrix} v_{p_1}(v_1) & v_{p_1}(v_2) & \dots & v_{p_1}(v_l) \\ v_{p_2}(v_1) & v_{p_2}(v_2) & \dots & v_{p_2}(v_l) \\ \vdots & \vdots & \ddots & \vdots \\ v_{p_k}(v_1) & v_{p_k}(v_2) & \dots & v_{p_k}(v_l) \end{pmatrix} \mod 2$$

Quadratic Sieve Matrix

$x_0 + r_i$	$v_i = (x_0 + r_i)^2 \bmod n$	Prime Factorization of v_i
265	-17238	$(-1)^2 \cdot 2 \cdot 3 \cdot 13^2 \cdot 17$
$\vdots$	$\vdots$	$\vdots$
278	-10179	$(-1) \cdot 3^3 \cdot 13 \cdot 29$
$\vdots$	$\vdots$	$\vdots$
$295 = \lfloor \sqrt{n} \rfloor$		
296	153	$3^2 \cdot 17$
$\vdots$	$\vdots$	$\vdots$
299	1938	$2 \cdot 3 \cdot 17 \cdot 19$
$\vdots$	$\vdots$	$\vdots$
307	6786	$2 \cdot 3^2 \cdot 13 \cdot 29$
$\vdots$	$\vdots$	$\vdots$
316	12393	$3^6 \cdot 17$

Table 1: Quadratic sieve for $n = 87463$ and search for 29-smooth square numbers

$$\begin{pmatrix} 2 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 3 & 2 & 1 & 2 & 6 \\ 2 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \end{pmatrix} \mod 2$$

Quadratic sieve matrix for the example $n = 87463$ with 29-smooth numbers. Rows are associated to the exponents of the factor base $-1, 2, 3, 13, 17, 29 \pmod 2$ appearing in the factorizations of these numbers.

In this example a kernel of the quadratic sieve matrix is $(1, 1, 1, 0, 1, 0)^\top$. So we compute the common divisor of

$$265 \cdot 278 \cdot 296 \cdot 307 \pm 2 \cdot 3^4 \cdot 13^2 \cdot 17 \cdot 29$$

and $n = 87463$.

Here is a short heuristic of why this works (well): Suppose we have l of the chosen $v_i's$, so

$$x \pm y = \underbrace{\prod_{i=1}^{l} (x_0 + r_i)}_{\sim n^{l/2}} \pm \underbrace{(\prod_{i=1}^{l} v_i)^{1/2}}_{\ll n^{l/2}}$$

and our only concern is when $\gcd(x \pm y, n) = n$ (both cannot be 1 because we know n divides $x^2 - y^2$). But the chances of this happening is almost like

$$\psi(n^{l/2}, n)/n^{l/2} \sim (l/2)^{-l/2}$$

which is rather low for l high enough.

There are other variations to the above method that we will not go too deep into (e.g. one chooses factor bases instead of B-smooth numbers, one uses other quadratic polynomial instead of $X^2 - n$ etc.).

Method 3 (Pollard's $(p-1)$-Method): The Pollard $(p-1)$-method takes advantage of numbers $n \in \mathbb{N}$ that are divisible by a prime p such that $p-1$ is B-power-smooth for some small positive $B \in \mathbb{N}$. Suppose this is the case and let $\beta(B) := \text{lcm}(2, 3, \ldots, B)$ and assume that for an $a \in \mathbb{N}$ with $(a, n) = 1$ we have $a^{\beta(B)} \neq 1 \bmod n$, then $\gcd(a^{\beta(B)} - 1, n)$ is a non-trivial divisor of n. In fact, this works best if $q-1$ is not B-power-smooth for at least one prime divisor q of n that is different from p. We can easily prove this in an exercise ...

Exercise 2. Let $n \in 2\mathbb{N} + 1$ be a composite number and $p, q \in P$ be such that $p \neq q$ and $pq \mid n$. Let $B \in \mathbb{R}$ such that $B > 0$ such that $p-1$ is B-power-smooth but $q-1$ is not B-power-smooth Prove that

$$\#\{a \bmod n : a^{\beta(B)} = 1 \bmod n\}/\phi(n) \leq \frac{1}{2}$$

This can be converted into an algorithm (given n and B)

Algorithm 2: Pollard $(p-1)$ Factorization Algorithm

Input: A composite odd number $n \in \mathbb{N}$, $B \in \mathbb{N}$ with $B \sim n^{1/4}$, `trials`$\in \mathbb{N}$
Output: A proper non-trivial integer factor of n

```
Compute β = lcm(2, 3, ..., B)
while i ≤ trials do
    Choose a random a in {2, ..., n − 1}
    if gcd(a, n) ≠ 1 then
        return gcd(a, n)
    end
    Set b = a^β mod n
    if gcd(b − 1, n) ∉ {1, n} then
        return gcd(b − 1, n)
    end
    Set i=i+1
end
```

One immediately may notice that this algorithm may face some problem. Firstly we are limited by the choice of B. So if neither of the prime factors of n are B-power-smooth, then it is unlikely that the algorithm returns a factor for n. We illustrate this by some examples

Example 2.

(a) Consider $n = 59 \cdot 101$. So $58 = 2 \cdot 29$ and $100 = 2^2 \cdot 5^2$. But neither 100 nor 58 is 20-power-smooth (100 is however 20-smooth). Let us compute the possibility of the Pollard-$(p-1)$ method to be successful with these inputs (i.e. n and B). We consider the smallest prime factor of n, $p = 59$. Let now $\beta = \text{lcm}(1, \ldots, 20)$ then we want to count

$$\{a \bmod n \, : \, a^\beta = 1 \bmod p \text{ and } a^\beta \neq 1 \bmod n\}$$

and take it's ratio with $\phi(n) = 58 \cdot 100$. Using Chinese Remainder Theorem and the Lemma in Lecture 02, the count is

$$\left(\frac{n}{p} - 1\right) \gcd(\beta, p-1) - \gcd\left(\beta, \frac{n}{p} - 1\right) \gcd(\beta, p-1) = 100 \gcd(\beta, 58) - \gcd(\beta, 100) \gcd(\beta, 58)$$
$$= 100 \cdot 2 - 2^3 \cdot 5 = 160$$

So the chances of getting a success using Pollard-$(p-1)$ for $B = 20$ and $n = 59$ is $\frac{160}{5800} \sim 3\%$ which is rather small. And this is explained by the fact that neither of the prime factors p of n satisfy the requirement that $p - 1$ is 20-smooth.

(b) Suppose now that we choose a case where B is rather high. Then you get the possibility that all prime factors p of n satisfy the condition that $p - 1$ is B-smooth. But this may results in $a^\beta = 1 \bmod n$ for all $a \in \mathbb{N}$ such that $(a, n) = 1$. For instance if n is square free, then $\phi(n) = \prod_{p|n}(p-1)$ and there is a high chance that this divides $\beta = \text{lcm}(1, \ldots, B)$ which implies that $a^\beta = 1 \bmod n$ for any $a \bmod n \in (\mathbb{Z}/n)^*$. As an example consider $n = 23 \cdot 89$ and suppose we chose $B = 20$. We know that both 22 and 88 are 20- power-smooth and both of them will divide $\beta = \text{lcm}(1, \ldots, 20)$. Let us consider $p = 23$. The chances of success of Pollard-$(p-1)$. We count all such $a \bmod n$ that $a^\beta = 1 \bmod p$ but $a^\beta \neq 1 \bmod n$

$$88 \gcd(\beta, 22) - \gcd(\beta, 22) \gcd(\beta, 88) = 0$$

In fact, this is the case for any $B \geq 20$.

(c) Consider again $n = 23 \cdot 89$. We can get the factorization using Pollard-$(p-1)$ if we choose $B = 8$. We cannot really choose a B such that 22 is B-power-smooth but 88 is not B-power-smooth. So we are practically forced in the situation of the first example.

But by a lucky chance we can choose $B = 8$ and compute $\beta = \text{lcm}(2, \ldots, 8)$ such that for $a = 12$ we get $a^\beta = 1 \bmod 23$ and $a^\beta \neq 1 \bmod 22$. Similar to the first example we check and see that the chances of choosing such an a is not very high.

There is an improvement to prime facorization that is roughly based on the idea of Pollard-$(p-1)$ but instead of relying on the group $\mathbb{F}_p^*$ for a prime factor p of n, it relies on a family of group whose order are within some interval with center p.

Method 4 (Lenstra's Elliptic Curve Method): We cannot describe this method in one sitting (we need to first get accustomed with *elliptic curves*. But we can give a general idea of it.

The Polard's $(p-1)$-method and some other factorization algorithms (e.g. $(p+1)$-factorization method) starts with the following idea:

(i) We have a sequence of group $\{G_i\}_{i \in \mathbb{N}}$, in the $(p-1)$-method $G_i = (\mathbb{Z}/i)^*$

(ii) For an $n \in \mathbb{N}$ that is composite and $p \in \mathbb{P}$ such that $p \mid n$ there is a non-injective group homomorphism
$$\phi_{n,p} : G_n \longrightarrow G_p$$
in the $(p-1)$-method this is given by the natural map $a \bmod n \mapsto a \bmod p$.

(iii) Getting a random element in G_n is easy even if G_n is not entirely understood.

(iv) If for the above, a non-trivial element $x \in \ker(\phi_{n,p})$ is found then we can obtain a non-trivial divisor of n

Even one does not know $p \in \mathbb{P}$ in advance, one hopes that we can find a small number $B \in \mathbb{N}$ such that

- $\#G_p$ is B-smooth or B-power-smooth.
- we can conclude that for a random $x \in G_n$ one has $\phi_{n,p}(x)^{\beta(B)} = 1$ in G_p and hope that $x^{\beta(B)} \neq 1$ in G_n

As we have seen in the example for Pollard's method, the above may fail. Especially if we really insist on $\#G_p$ being B-power-smooth (for our example, if we are lucky, we may still succeed with Pollard's method but we should not insist on B-power- smoothness of $\#G_p$).

In Lenstra's Elliptic Curve's method, for each prime divisor p of n we can actually produce a family of groups $\{G_{p,j}\}_{j\in J}$ (including family of homomorphism mapping G_n to $G_{p,j}$) and one can make statistical statements on how high the chances are that one group in the family will give us a B-power-smooth number for which we can succeed in finding a non-trivial element of the kernel of the corresponding homomorphism.

In other words, the chances of a succesful factorization with ECM are much higher. But to understand this method and why it works, we need to understand elliptic curves.

So we will postpone describing this method until we actually know more about elliptic curves.